DOCUMENT RESUME

ED 411 263 TM 027 208

AUTHOR Auchter, Joan E.; Stansfield, Charles
TITLE Linking Tests across Languages: Focus on the Translation and Adaptation Process.
PUB DATE 1997-03-00
NOTE 22p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (Chicago, IL, March 25-27, 1997).
PUB TYPE Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Bilingual Students; English; *Equated Scores; *Equivalency Tests; High School Graduates; High Schools; *Spanish; Spanish Speaking; Test Format; Test Items; *Test Use; *Translation
IDENTIFIERS *Linking Metrics

ABSTRACT

Is it possible to translate an existing test from the source language to a target language and then to empirically link the translated test to the source language standard? Some tests are amenable to translation of their content and some are not. Within tests, some items may be translatable and others may not. Some items may measure meaningful content when translated, and others may not. Within items, some options may be translatable, and others may not. This paper focuses on the rigorous method used to translate and adapt from English to Spanish three equated forms of the Tests of General Educational Development (GED tests) to ensure that all items are valid and that the translated instrument measures a comparable construct. Most specifically, the paper addresses issues related to translating items on the language subtest. The Spanish-language GED tests were normed using only graduated high school seniors in Puerto Rico. In addition, the paper outlines a linking design that introduces a procedure for screening biliterate students for linguistic equivalency across languages before including them in the linking sample. This design will be tested in May 1997 when the test is administered to a large sample of biliterate Hispanic high school seniors. (Contains 2 tables and 12 references.) (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 411 263

Linking Tests Across Languages: Focus on the Translation and Adaptation Process

Joan E. Auchter, GED Testing Service of the American Council on Education

and

Charles Stansfield, Second Language Testing, Inc.

Abstract

Is it possible to translate an existing test from the source to a target language and then to empirically link the translated test to the source language standard? Some tests are amenable to translation of their content and some are not. Within tests, some items may be translatable and others may not. Some items may measure meaningful content when translated, and others may not. Within items, some options may be translatable and others may not. This paper focuses on the rigorous method used to translate and adapt from English to Spanish three equated forms of the Tests of General Educational Development (GED Tests) to ensure that all items are valid and that the translated instrument measures a comparable construct. Most specifically, the paper addresses issues related to translating items on the language subtest. In addition, the paper outlines a linking design that introduces a procedure for screening biliterate students for linguistic equivalency across languages before including them in the linking sample.

Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago, Illinois, March 1997.

TM027208

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

Joan Auchter

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

BEST COPY AVAILABLE

Linking Tests Across Languages: Focus on the Translation Process

Can a forward-translation of a test from English to Spanish result in a test instrument that removes the language barrier while maintaining the same content and standards? The purpose of the GED Testing Service Spanish Test Development Project is to develop, based on an analysis of the issues, Spanish-language GED Tests that correspond to the U.S. curriculum norm group, so that the GED candidates' Spanish-language scores are comparable to the scores of candidates who take the English-language GED Tests.

BACKGROUND

Description of the English-language GED Tests. The GED Tests are designed to provide an opportunity for persons who have not graduated from high school to earn a high school level diploma that is recognized by both institutions of higher education and by employers. Administered in all fifty states and the territories in the United States, and in most Canadian provinces and territories, almost 800,000 people take the GED Tests annually. Approximately one of seven high school diplomas issued annually in the United States is a GED diploma.

The third generation of GED Tests, introduced in 1988, is a 5-test battery that requires 7 hours and 45 minutes of test administration time. A GED candidate earns a GED diploma only after passing all five tests. The official titles of the five separate subject tests, and their time limits, are as follows: Test 1: Writing Skills (120 minutes), Test 2: Social Studies (85 minutes), Test 3: Science (95 minutes), Test 4: Interpreting Literature and the Arts (65 minutes), and Test 5: Mathematics (90 minutes). The Writing Skills Test has two parts: Part I is made up of multiple choice questions and Part II is a direct writing assessment (essay); the other four tests contain only multiple-choice questions.

To allow GED candidates the opportunity to demonstrate achievement comparable to that of high school graduates, the tests are based on two foundations: 1) test content that conforms as closely as possible to the core academic curricula of the U. S. high schools, and 2) score scales based on periodic norming of the GED Tests on a nationally representative sample of graduating U.S. high school seniors. This norming process allows the passing standards for the

GED Tests to be referenced to the actual performance of those who graduate via the traditional route. The minimum passing score is set so that approximately 66% of graduating U.S. high school seniors would pass the test battery and 34% would fail.

Description of the Spanish-language GED Tests. The Spanish-language GED Tests were originally developed to provide adults in Puerto Rico who had not graduated from high school an opportunity to earn a GED diploma comparable to the diploma awarded by the high schools in Puerto Rico. (Spanish is the primary language of instruction in Puerto Rican high schools.) The 1988 revised Spanish-language GED Tests, introduced with the revised English-language tests, include content changes recommended by the Puerto Rican curricular experts and content specialists involved in development of the tests. For a direct comparison of the tests, see *The Tests of General Educational Development Technical Manual, First Edition,* (Auchter, 1993).

The Spanish-language GED Tests were normed using only graduating high school seniors in Puerto Rico. Due to the increasing number of Spanish -speaking adults throughout the U.S. without a high school diploma, many states began offering the Spanish-language GED Tests to their Spanish-speaking GED candidates. Currently, the Spanish-language GED Tests are taken more often in the continental U.S. than in Puerto Rico. In 1995, about 26,500 Spanish GED Tests were administered in the mainland U.S. and about 14,600 were administered in Puerto Rico (GEDTS, 1996).

STATEMENT OF THE PROBLEM

The Spanish-language GED Tests do well what they were developed and normed to do: provide an opportunity for adults in Puerto Rico to earn a GED diploma comparable to the diplomas awarded by high schools in Puerto Rico. The use of the Spanish-language tests outside of Puerto Rico has been criticized because some states offer the same high school level credential, regardless of the particular language version of the GED Tests taken. This use is considered inappropriate because the content of the two test versions varies, the two language versions of the tests are normed on different populations, and the score scales are not linked. Thus, it is possible that different levels of ability are required to obtain the same GED score, and therefore, the credential. As a result of these concerns, the GED Advisory Committee and the Commission on Educational Credit and Credentials, the governing board of the GED Testing

Service, required the GED Testing Service to produce a new Spanish-language version of the GED Tests that will be comparable in content and difficulty to the English-language GED Tests.

METHODS

To determine if the goals of this project are obtainable, the GED Testing Service first conducted a preliminary analysis of the translatability of the GED Tests, and then commissioned three feasibility panels to explore the technical issues involved in linking the two language versions of the GED test battery.

Translatability Study. Prior to convening the three feasibility panels, GEDTS contracted with an outside firm to conduct a preliminary translation study evaluating the feasibility of a forward-translation of the English-language GED Tests into Spanish. The purpose of the study was to discern whether or not test items were amenable to translation to Spanish. The results of the analyses, conducted under a very stringent method, reveal that current forms of each test could be directly translated with minor modifications. Complete analyses are included in the GED *Direct Translation Feasibility Study (Colberg, 1993).*

Feasibility Panels. After it was clear that the language and content of the test could be translated, GEDTS commissioned the formation of a series of feasibility panels to explore the technical issues involved. The first panel, the Psychometric Feasibility Panel (PFP), was convened in October 1993 to investigate the feasibility of linking the English and Spanish-language versions of the GED Tests. The PFP consisted of the following four prominent psychometricians who were selected for consultation based on their experience in psychometrics and test equating: Linda L. Cook of the Educational Testing Service, Ronald K. Hambleton of the University of Massachusetts at Amherst, David Thissen of the University of North Carolina at Chapel Hill, and Howard Wainer of the Educational Testing Service. This panel discussed seven tentative linking options, evaluated IRT and anchor item linking procedures, and evaluated the utility of matching in linking and made a number of recommendations concerning linking options. The committee's deliberations were presented as a document, *Linking the English-language and*

Spanish-language Versions of the Tests of General Educational Development: Psychometric Feasibility Study, (Sireci, 1994).

A second panel, the Linguistic Feasibility Panel (LFP) which met in March 1994, considered translatability and other issues concerning the development of the new Spanish tests, and also made a number of recommendations. The LFP panel members, selected for consultation based on their expertise in linguistics and experience with language testing, included: Brunilda deLeon of the University of Massachusetts, Pardee Lowe, Jr. of the U.S. Government Language School, Cecelia Rosenblum of the Educational Testing Service, Ramon Santiago of Lehman college, Charles Stansfield of Second Language Testing, Inc., and Gillian Whalen, Spanish Linguist. The LFP panel agreed with the translatability study's conclusion that most items can be translated. However, it noted that some items in the Writing Skills Test would not test meaningful or challenging content in Spanish. Particularly important were the panel's translation process design and its recommendations concerning the feasibility of translating the Writing Skills Test into Spanish. The panel noted that the Spanish-language version of the Writing Skills Test would tap a different but related construct. The two different constructs would be writing skills in English and writing skills in Spanish. While the constructs both relate to writing skills in the native language, they are nonetheless quite different abilities. Thus, it may not be possible to report the score on the Writing Skills Test in the same way it would be possible to report the score on a translated test of a subject other than language. Clearly, the development of a Spanish-language version of the Writing Skills Test would involve adaptation, in other words, a significant modification of the instrument. The LFP deliberations are summarized in the *Development of Revised Spanish-language Versions of the Tests of General Educational Development: Linguistic Feasibility Study* (Auchter, 1994).

The third panel, the Combined Feasibility Panel (CFP), met in October 1995 to consider the findings of both previous panels, and to make more detailed recommendations as to how to proceed with both the translation and the linking of the two language versions. The CFP reviewed desirable background characteristics of translators, sampling issues, linking designs, and procedures to check for comparable standards. The following CFP members were selected based on their expertise in linguistics, second language testing, cross-lingual assessment, and psychometric methods: Ronald Hambleton of the University of Massachusetts, Pardee Lowe, Jr. of the U.S. Government Language School, Maria Pennoch-Roman and Alicia Perez Schmitt of the Educational Testing Service, and Charles Stansfield of Second Language Testing, Inc. The CFP developed a rigorous translation model based on the CFP report and the *Guidelines for Adapting Educational and Psychological Tests* from the International Test Commission (ITC, in

press; largely summarized by Hambleton, 1994). This translation model is reported in the *Options for the Development and Linking of New Spanish-Language Versions of the GED Tests* (Auchter and Stansfield, 1996). In addition, this panel reviewed and approved the *Action Plan for the Development and Linking of New Spanish-Language Versions of the GED Tests* (GEDTS, 1996) that is discussed later. While the CFP's focus was the entire test, approximately half the panel's deliberations focused on the Writing Skills Test.

TRANSLATION PROCESS

The translation of a test into another language is an important task. It is assumed by test score users that the translated items are equivalent in meaning and difficulty to the original version in English. This equivalence reinforces the claim for score comparability. If the translation is accurate, then the examinee will not be affected (assisted or disadvantaged) by the quality of the translation. Thus, the examinee's response to each item will reflect the ability to respond in his or her native language to the exact same item that was administered in English to English proficient students.

Similarly, a translation must be expressed in natural language, or in language that is as natural as the language used in the English original. If a translation is too literal, it will read like a translation as opposed to an authentic document in the target language. This lack of naturalness in the wording of the item often results in poor quality items which, generally, are more difficult. Furthermore, it is generally know that research on item bias shows it is often the least able examinee who is most disadvantaged by awkwardly worded items. If a translation is too literal, then the meaning of the original item will be distorted because a critical distinction in the original may be simplified or not carried over to the translated tests. Normally, a distortion in meaning makes it less probable that the examinee will perform well on the item. The resulting loss of information makes the item harder to answer correctly (Stansfield, 1996). For additional information on the translation of tests, see (Hambleton, 1994).

Our presentation of the translation process is divided into two broad groupings: general guidelines which apply to all five subtests and issues which are specific to individual subject-area subtests. The general guidelines are addressed first and the issues specific to particular subject-area subtests follow.

General Guidelines in the Translation Process. While there are issues specific to each of the five subject-area tests, the following steps applied to all five tests.

Step One: Selecting Three Forms Most Appropriate for Translation

Seven equated operational forms of the Science, Social Studies, Mathematics, Interpreting Literature and the Arts, and Writing Skills Tests were inspected and compared to select the three forms best suited to translation. The two reviewers were native Spanish-speakers who have worked on the GED Tests for a number of years. In addition to a native command of Spanish, both reviewers have extensive test development experience. The evaluation method was designed to ensure the possibility of evaluating test forms with sufficient thoroughness to permit the identification of those test forms which exhibit the greatest potential for parallelism between the English-language original and the Spanish-language translation. From the evaluation, it appeared that all items on the three forms in Social Studies, Science, Interpreting Literature and the Arts, and Mathematics could be successfully translated. While most items on the Writing Skills Test could be successfully translated, it wasn't clear if the translated items would measure meaningful skills and knowledge. This issue will be addressed in the Writing Skills specific issues section.

Step Two: Selecting Translators

After the test forms were identified, the primary translators were selected. General requirements for translators, as defined by the LFP and CFP, include: 1) American Translation Association (ATA) or equivalent certification, 2) near-native reading and writing skills in English, the source or donor language, 3) educated native-writing skills in Spanish, the target or receptor language, and 4) congruity judgment, which is the ability to judge the equivalence of the original and translated text in terms of their meaning, style, and rhetorical structure, 5) experience in the test development process - ideally, experience as an item writer; translators familiar with the mechanics and rules of item writing should be attentive to grammatical clues, clang associations, the length of the correct answer, and the homogeneity and parallelism of the answer choices, 6) appropriate academic and subject specialization; different translators were selected according to their area of specialization. Interpreting Literature and the Arts required a specialized literary translator, someone who is also skilled as a creative writer.

Step Three: Translator Orientation

Because tests represent a different kind of text than translators routinely handle, the proper and detailed orientation of translators is especially important. Prior to beginning the translation, translators were given basic information on the GED Testing Program and the test population. Translators also were given a copy of the test specifications for the tests they were translating, the *Technical Manual for the Tests of General Educational Development*, the *Item Writer's Manual* furnished to English-language GEDTS item writers, and the *Standards for Adapting Instruments and Establishing Score Equivalence* (1993).

Translators were given a copy of the English-language versions of the tests, including graphics, and were requested to provide the translation of all text, including titles and footers. The importance of translating each message or proposition within each test task was emphasized. Translators also were coached to be aware of dialect and syntax issues. Since GED examinees are expected from all Spanish-speaking countries, the translators were advised to make a conscientious effort to use language that is not biased toward the peculiarities of any particular national speech. The language should be as clear to a person of Argentine roots as to one of Mexican or Spanish heritage. Terms that vary across dialects also pose a considerable problem that translators must address. In this case, it was decided to consider all possible variants of a word or phrase, and then to look for the variant that is most neutral or most widely understood across the Spanish-speaking world. An example is the word for car in Spanish. Depending on dialect, a speaker might commonly use *coche, carro, or maquina*. Each of these words could mean something different to speakers of the other dialect. Yet the word *automovil* would be understood by speakers of all these dialects.

Translations from English into Spanish all too frequently retain the use of the passive voice when it would be more germane to a Spanish-language text to use the active voice. The result is an anglicized text that is structurally inappropriate, and, hence, occasionally unintelligible. Finally, translators were requested to compile a file of comments concerning any items in the tests that they viewed as problematic, or any portions of the tests that posed the problems for translation. These comments were sent to reviewers for their consideration during the review process.

Step Four: Initial Forward-Translation

The LFP and PFP agreed that forward-translation followed by reviews and revisions was appropriate. Each form of each subject-area test was translated from the source to the target language by the principal translator, who thereafter revised and refined the initial translation in successive iterations.

Step Five: Initial Review

This initial translation was reviewed and further refined by the principal reviewer who was asked to compare the translation with the English-language version of the test. The reviewer created a list of specific concerns and suggested revisions. This list was returned to the project manager.

Step Six: Project Manager Review

Each test was returned to one of the two project managers who have extensive experience in test development and translation. After reading the translation and review, the project manager discussed the issues with the reviewer and then the primary translator. The primary translator made the indicated corrections, along with a list of suggestions not implemented, with a justification for each.

Step Seven: Secondary Review

The final translation was then reviewed by two secondary reviewers who were sensitive to language, dialect, regional and cultural differences. These reviewers reviewed the translations for linguistic accessibility and equivalence in meaning. The secondary reviewers also read the translation for the naturalness of expression in Spanish, making revisions as appropriate so that the text would not appear to be a translation. These problems and suggested revisions were returned to the project manager, who returned them to the principal translator. Again, the principal translator either made the revisions or documented why the revisions were not made.

Step Eight: Principal Translator Key Verification

With the translation in final form, the principal translator read each test against the English-language original and verified the correct response for each test item. This step added additional verification to the translation and corroborated the viability of all keyed responses and all incorrect response choices, thus corroborating the preservation and integrity of the instrument.

Step Nine: Translation Documentation

Because the quality of a translation is critical to the reliability, validity, and score comparability of a test, it is necessary to document the process that was followed to translate a test. Each project manager was required to document the process and efforts that were made to ensure the quality of the translation of each form of each of the GED subject area tests, as well as the professional qualifications of the translators who performed the translations and the reviews.

Issues Specific to Individual Subject-Area Subtests. In addition to the general steps described above, there are several issues which are relevant to the documentation of the translated versions of the specific subtests. In addition, there is an adaptation process required for the Writing Skills Test, which will be described following this section.

Test 2: Social Studies

While the Social Studies Test includes fewer technical terms than the Science Test, due to its length and content it poses significant translation challenges. Since much of the content is U.S. specific, it can pose challenges for rendering to another language. For example, terms such as "freedom rider" have no equivalent in Spanish, resulting in the need to paraphrase or define. Each primary translator made a detailed list of these translation issues. There were no issues in any of the three translated forms that required major changes or items or item substitutions. Therefore, all items on the Spanish-language Social Studies Test are direct translations of the English-language items and can potentially be considered as anchor items.

Test 3: Science

A viable, faithful translation depends on the translator having a strong science background and knowledge of how science concepts are expressed in the target language because of the technical terms included in the text. There are many classifications in the taxonomies of the animal and plant kingdom that escape the memory of the strictest specialists

(Colberg, 1996). The translator must have a strong enough command of science to replace more technical terms with less technical but more appropriate terms that refer to the same scientific phenomenon. There were no major changes in items or item substitutions on the three Science Test forms. Therefore, all items on the Spanish-language Science Test are direct translations of the English-language items and can potentially be considered as anchor items.

Test Four: Interpreting Literature and the Arts

The method of translation followed with the Interpreting Literature and the Arts Test was essentially the same as the method described above, although the translation of poems and dramatic excerpts required greater reliance on the interaction among translators. Prior to the translation of the literary selections, an extensive search was conducted to ascertain whether or not a translation was available. If more than one translation was available (as would be the case with the major authors), these translations were evaluated to determine which was the best one. In those cases where the published translations were available, these translations were analyzed and compared sentence by sentence with the English-language original in order to ensure that the translator had been totally faithful to the original text. Those selections for which published translations were not found were translated with great care to preserve the meaning of the text without sacrificing its literary value. In the work on the Interpreting Literature and the Arts Test, no items were identified by the translator as unsuitable in the translated version. Therefore, all items on all three Interpreting Literature and the Arts Tests are direct translations of the English-language items and could be considered as anchor items.

Test Five: Mathematics

There were no translation issues with the Mathematics Test. The most contentious issue to arise was how to translate the term "right triangle." Since two different terms are used in different parts of the Hispanic world, the two project managers decided to use one term, but insert the other in parentheses as a gloss.

WRITING SKILLS TEST ADAPTATION PROCESS

While the method of translation followed with the Writing Skills Test includes the same steps as the method described above, this test required several additional steps. A Translatability Feasibility Study (Colberg, 1993) indicated that it is possible to translate directly all Writing Skills Test items except those that measure the use of possessives or phonetically based

errors. The major problem is in the meaningfulness of the items and tasks resulting form the translations.

The Writing Skills Test is a two part test. The multiple-choice portion consists of 55 items that test knowledge of the structure and conventions of standard written English. The essay portion consists of a 45 minute writing sample based on a specific prompt. A dozen English prompts that functioned well during pretesting were reviewed for their accessibility to a Hispanic examinee population. About half the prompts were judged to be about equally accessible to Hispanics. These prompts were translated to Spanish and are used with the current Spanish-language versions of the Writing Skills Test.

The writing samples are rated using a six point holistic scoring guide. The guide had previously been translated to Spanish for use with the current Spanish-language GED Tests developed for Puerto Rico. Since it was first translated, it has been used many times by the GEDTS pool of bilingual readers, and over the years, several minor revisions and additions have been made to better reflect the linguistic features of standard written Spanish. The prompts and the scoring guide are usable as part of the new Spanish language version of Test 1.

Multiple-choice items in Test 1 items are based on a stimulus text which contains various kinds of errors in writing that have been deliberately introduced into the text. For each sentence containing an error, the examinee must choose from five options the one that would make the sentence correct. Many of the translated items are valid measures of knowledge of the structure and conventions of formal written Spanish. For example, items that test subject-verb agreement or coordination of tenses across clauses in English will normally test similarly valid knowledge and skills when translated to Spanish.

However, other items pose a problem when translated. Items that test spelling in English are often less valid, because the particular word that contains the spelling problem does not pose a spelling problem when translated into Spanish. An example is knowledge of the difference between "there" and "their." If one of these were tested and the other used as the basis for distractors on the Spanish exam, probably no examinee would confuse the usage of the three Spanish equivalents, "hay," "su," and "allí." Similarly, for the distinction between the usage of "would" and "wood," examinees would never confuse "madera" with "hubiera." So it was clear from the results of the translatability study that some translated items would have to be modified

or even completely replaced. Finally, written Spanish contains diacritic marks not found in written English. These are just a few of the ways in which Spanish differs from English.

Initial translation. In order to deal with these issues, GEDTS initiated an iterative test adaptation process, which has recently been completed. The first step was the selection of an appropriate translator. The International Test Commission's *Guidelines for Adapting Educational and Psychological Tests* (Hambleton, 1994) call for the use of translators who know both languages and cultures, the content of the discipline in which the translation will be done, and have experience in item writing and the test development process. The translator selected to translate Test 1 was a native speaker of standard Latin American Spanish who has lived in the U.S. for 25 years, is an experienced professional translator accredited by the American Translators Association, has taught Spanish at two universities before becoming a professional translator and interpreter, and has experience as an item writer and developer of tests for the selection of translators by U.S. Government agencies.

The lead translator having been identified, the next step was the selection of the forms of the test to be translated. First, GEDTS staff selected seven forms of the test that reflected the latest revisions to the test specifications and that had shown good psychometric characteristics through pretesting and operational administration. These seven forms were then reviewed by the translator who did the translatability study mentioned earlier. She was tasked with identifying the three forms most suitable in content and language for translation. These forms were then translated to Spanish by the primary translator selected for this project. The primary translator was instructed to translate all items whose content was translatable and to modify items or options that were not using the same stimulus sentences that were used in the English original.

Upon completion of this work, the translation was reviewed by another experienced translator who served as a primary reviewer. Suggested corrections and revisions were either implemented or a written explanation as to why they were not implemented was provided in a separate document. In addition, the primary translator created a document that discussed any difficulties surrounding the translation of each item and option. The document also classified the translated item according to the content specifications for GED Test 1 and each item was classified as being the same as the English original or a modification.

Charge to the Advisory Committee. The GED Testing Service then convened a four person national Test Advisory Committee for the Spanish-language version of Test 1. The role of the committee was to specify the degree of adaptation needed through recommendations for an

adapted set of specifications and to approve all items that would appear on the test. All members had extensive experience teaching Spanish to native speakers in this country at either the high school, community college, or college level and all were experienced in the test development process. Three of the four had previously served as full-time employees of another test publisher and three of the four were native speakers of Spanish. The Advisory Committee members included Raul Rodriguez, Rodney Rodriguez, Carmen Salazar and Charles Stansfield.

The Advisory Committee members were sent background information on the GED program, background information on the deliberations of the three panels that had been previously convened to discuss the translation/adaptation of the GED tests, the English original and translations of each form of Test 1, a content analysis for each form showing how each item is classified in the specifications and its p value, and the *Test 1 Item Writer's Manual* in English (which contains the test specifications). They were instructed to review the *Item Writer's Manual*, the English-language test forms, and the Spanish-language translations, and to make detailed notes and comments concerning revisions that should be made in the translated items. They were also instructed to identify any items that would be inappropriate for a Hispanic examinee population.

Recommendations of the Committee. The Advisory Committee met for a total of 18 hours during two days in October, 1996. The first day of the meeting was devoted to reaching consensus on the wording of each stimulus text and each item on one form of the test. The discussions were based on the detailed comments that committee members had written on the tests they had been sent. Members also classified each item as being either a direct translation from English, a minor change from English (involving only one option), or a major change from English (involving two or more options). For the one form examined, it was found that 67% of the items initially provided by the translator/item writer were direct translations of the English original, 22% involved a minor change, and 11% a major change from the English original.

The second day of the meeting was devoted to a discussion of how well the English-language test specifications applied to the Spanish-language version. The English-language test specifications are presented, along with examples, in the Item Writer's Manual, which was used as a point of departure. Committee members began by reviewing the three item types on the

Writing Skills Test. They pointed out that often one item type seemed awkward in Spanish and suggested modifications to alleviate the problem.

Next, they discussed the content categories included in the specifications for the English-language version. As they progressed through each category, they indicated whether it applied to Spanish and to the same degree, i.e., whether a content category was more or less important in Spanish than in English. They identified categories that could be deleted from the Spanish version of the Manual and added several new categories. Some of the new categories are: a) Gender and Number Agreement of all types, b) Prepositions, c) Other Troublesome Words, and d) Accents and Other Diacritic Marks. The review of content categories produced a shift in the general content distribution as follows:

Writing Skills Test Content Categories

CONTENT CATEGORY	ENGLISH	SPANISH
SENTENCE STRUCTURE	25%	30%
U.S.AGE	45%	40%
MECHANICS	30%	30%

Another important issue dealt with by the committee was the kind of examinee for whom the item is being written. Examinees taking this test may be either monolingual or bilingual. Monolinguals make mistakes in writing due to an imperfect command of written Spanish. Bilinguals make these errors, too, plus errors that result from interference from English. It was decided to write items that would test common confusions of the monolingual only. Otherwise, bilinguals would be disadvantaged, since they would find both types of items challenging, while monolinguals would only find one type of item challenging.

Finally, the committee discussed the issue of different oral Spanish dialects. It was decided that the testing of nonstandard but widely used verb forms such as "haiga" for "haya" was appropriate. On the other hand, it was decided that the spelling problems tested should not be dialect based, since this would disadvantage the speakers of those dialects. For example, only Puerto Ricans might confuse *l* and *r* because of their substitution in Puerto Rican spoken dialect. Because Spanish has some special standard dialects, forms of those dialects should not be tested, since that would advantage speakers of those dialects. Thus, no verb forms associated with the voseo, an alternative conjugation used in Central America and southern South America, should be tested; nor should be vosotros forms used in Spain. On the other

BEST COPY AVAILABLE

hand, errors associated with the language development process (child-like speech), such as "cabo" for "quepo," should be tested.

Detailed minutes of the Advisory Committeé meeting were prepared and sent to the committee for review, revision, and eventual approval. Subsequently, the minutes and copies of the detailed comments on each item were sent to the translator/item writer. These were used to revise the tests and to create new items to fit the new content categories. The primary translator then prepared a revision of the Item Writer's Manual for the Spanish Test. This was titled the *Item Writer's/Translator's Manual for the Spanish-Language Version of Test 1*. It includes the revised content specifications, new or modified content categories, and examples in Spanish of stems for each item type within each content category. It also establishes the exact wording of formulaic expressions that appear repeatedly in item stems and options. It includes a sample translation of a text, the associated sample items and discusses how to use existing text to test new content categories.

The Advisory Committee was very pleased with the additional revisions, the new items written, and with the Item Writer's/Translator's Manual. Naturally, however, they still had some additional revisions to suggest in the wording of the translated stimulus texts, and they identified a few items which they felt might have double keys. Their concerns have been accommodated, and the test is now finalized.

Analysis of changes in items. The final version of the test was then analyzed to determine the relationship of each item on the translated version to the original English version. This content analysis shows for each item, the item type used and the content category and subcategory tested. It also gives an analysis of the similarity of the translated version to the English original for each item and option.

Table 1 that follows depicts the results of the analysis for the total test for the three forms. Items are classified as:

- direct translations from English or same as English (SAE)–judged to be content valid in both languages,
- exhibiting a minor change (MC) involving only one distractor,
- exhibiting a big change (BC) involving changes on two or more distractors,

- testing content categories that are unique to the Spanish specifications (S), through use of the translated version of the original English stimulus sentence.

Cases where new stimulus sentences were used to construct new items are also identified, and they are subdivided by those that test Spanish specific content categories (New-S), and those that test content categories that also appear on the English test (New). Sometimes items of the latter type were constructed:

- to ensure content balance within or across the forms,
- because the wording of a translated item was awkward,
- because the translated item would clearly have been easier or more difficult in Spanish, or
- because the item subtype itself posed unique problems when converted to Spanish.

Table 1. Classification of Writing Skills Items in Adapted Spanish Form

	FORM AM					
	SAE	MC	BC	S	New-S	New
Number Items	26	13	9	6	1	
Percent	47%	24%	6%	11%	2%	
	FORM AQ					
	SAE	MC	BC	S	New-S	New
Number Items	26	1	11	10	5	1
Percent	47%	2%	20%	18%	9%	2%
	FORM AS					
	SAE	MC	BC	S	New-S	New
Number Items	27	11	13	3	1	
Percent	49%	20%	4%	5%	2%	

SAE - Same as English (dual content)
MC - Minor change (1 option changed)
BC - Big change (2 or more options changed)
S - Spanish specific content which would not apply to English
New-S - New stem with Spanish specific content
New - New stem with dual content

BEST COPY AVAILABLE

The results of the comparison of the English and Spanish versions shows that almost half of the original items were directly translatable to Spanish. About 20% of the items required that two or more distractors be changed, and another 20% of the translated stems resulted in changes that reflect Spanish only content categories. The remaining 10% of the items involve new stimulus sentences and options. One item required only a change of one distractor.

ESTABLISHING EMPIRICAL LINKS

During May of 1997, the test will be administered to a large sample of biliterate Hispanic high school seniors. Subjects in the sample will be selected based on the similarity of their performance on a screening test in both languages. The screening test will be Test 4 of the GED, which is a measure of reading comprehension involving passages from literature and the arts. Each examinee will receive a bilingual test booklet, with half of the items in English and half in Spanish. The order of presentation will be counterbalanced. Examinees whose score on each half differs by only two standard errors of measurement (3 raw score points) will be considered to be balanced biliterates. From this group, a sample will be selected which is as similar as possible to the distribution of ability within the sample of 12th grade students that was used to establish norms and the cut score for the GED in English in 1996. These students will then take different forms of Test 1 in both English and Spanish. The order of presentation will again be counterbalanced. Test 1 English item parameters from the biliterate sample will be will linearly transformed to the same scale as the 1996 standardization. Items on the Spanish version that have been identified as identical to those on the English version will be considered for use as an anchor test. The tentative anchor items on the Spanish test will be linearly transformed to the1996 standardization scale. Those with similar item parameters will then serve as an anchor to calibrate the scores on the Spanish-language version and link it to the score scale for the 1996 standardization of the English-language version. In this way, we expect to be able to assure score users that scores on the Spanish test reflect a degree of mastery of the construct that is comparable, within a specified error of measurement, to the equivalent scores on the English-language version of Test 1.

CONCLUSIONS

By summarizing the results of the GED test adaptation process, we can make the following observations:

- Guidelines exist that promote rigorous translation from a source to a target language. By stringently following these procedures, it is possible to minimize differences in item difficulty introduced when translating tests from one language to another.

The surprising observation is that translation is generally done long after the source language test is developed and standards are set. From our translation experiences, we have identified the following guidelines for future English language test development that would smooth the way for translation of new tests:

- Avoid stimuli that reference topics identified with American culture, such as baseball. For example, in baseball, which is not an international sport, there is no translation for "shortstop." A careful review of all stimulus should be done prior to developing items.

- Select literary texts, when possible, for which a translation already exists. By using texts that have published translations, the time spent translating would be eliminated.

- Add translation reviewers to the item and test review stages of test development. These reviewers can identify potential translation problems and suggest revisions during the test development stages.

While translation from one language to another may not result in tests that are equivalent in both languages, careful attention to translation issues during the English- language test development process and strict adherence to established translation guidelines can reduce the likelihood of introducing bias factors that can lead to differences in performance.

References

Auchter, J.C. & Stansfield, C.W. (1996). Action plan for the development and linking of new Spanish-language versions of the GED Tests. Washington, DC: GED Testing Service.

Auchter, J.C. & Stansfield, C.W. (1996). Development of revised Spanish-language versions of the tests of General Educational Development: Linguistic feasibility study. Washington, DC: GED Testing Service.

Baldwin, J. (1996). Who took the GED? Washington, DC: American Council on Education.

Colberg, M. (1993). Direct translation feasibility study. Washington, DC: GED Testing Service.

Colberg, M. (1993). Translation of the tests for General Educational Development: Documentation report by Logos, Inc. Washington, DC: GED Testing Service.

GED Testing Service. (1990). Item writer's manual, Test 1: Writing Skills. Washington, DC: Author.

GED Testing Service. (1997). Item writer/translator's manual for the Spanish-language version of Test 1: Written Expression. Washington, DC: Author

Hambleton, R. (1994). Guidelines for adapting educational and psychological tests: A progress report. European Journal of Psychological Assessment, 10, (3), 229-244.

Sireci, S.G. (1994). Linking the English-language and Spanish-language versions of the Tests of General Educational Development: Recommendations of the GED-STEP Psychometric Feasibility Panel. Washington, DC: GED Testing Service.

Sireci, S.G. (1997). Problems and issues in linking assessments across languages. Educational Measurement: Issues and Practices, 16, (1), 12-19, 29.

Stansfield, C.W. & Auchter, J.C. (1996). Options for the development and linking of new Spanish-language versions of the GED Tests: A report emanating from the deliberations of the Combined Feasibility Panel. Washington, DC: GED Testing Service.

Stansfield, C.W. (1996). Report on translation procedures employed by Second Language Testing, Inc. Washington, DC: GED Testing Service.